Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at http://about.jstor.org/participate-jstor/individuals/early-journal-content.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

THE SERVICE AND IMPORTANCE OF STATISTICS TO BIOLOGY.*

BY RAYMOND PEARL, PH.D., *Biologist, Maine Agricultural Experiment Station.*

The key note of anniversary celebrations is traditionally the contemplation of the achievements of the past. The American Statistical Association is today celebrating its seventy-fifth birthday. This for an individual is a very respectable old age, even in this day of lactic defenses against senility. In the life of a society, however, and especially of this Association, seventy-five years mark merely the beginning of life—the attainment of full youthful vigor—rather than the approaching dissolution of old age. Statistical science, which has been so ably promoted by this Association during the years of its existence, is still a growing and a developing science. By the discovery of new methods its power to solve problems is every day becoming greater, and every day new problems are being put before the statistician for solution by workers in other fields of science. Only a short time ago I was told by a distinguished biochemist that in his judgment many of the most fundamental problems relating to the chemical physiology of the animal body could only be solved by the application of statistical methods. Without these methods but little real progress on certain of these problems had been made. Great as have been the achievements of statistical science and of this Association in the past, one may well believe that they mark but the beginning, and will be far surpassed by new developments and applications of statistical methods of reasoning, in fields now either untouched or only just entered upon.

It is about one of these newer developments of statistical science that I have been asked to speak today. This is the application of statistical methods to the problems of biology. While relatively a new field, this branch of science has developed very rapidly during recent years. It has been given a special name, biometry, and a number of universities now offer courses in the subject. A special journal—*Biometrika*—devoted to this field is now in its ninth volume. This occasion marks, I think, the first formal recognition of the existence of this youngest of its daughter sciences by the parent science of statistics. It is a great pleasure to assist at this "coming out party," in which I shall essay something like the part usually played by the débutante's frock, endeavoring modestly to draw attention to and set off her inherent charms and graces.

---

* Paper read at the seventy-fifth anniversary meeting of the American Statistical Association, Boston, Mass., February 14, 1914.

The application of statistical methods to the study of biological problems other than those of anthropology may be said to have begun with the work of the late Sir Francis Galton. Galton was a born statistician. He tells in his "Memories" of the instinct, which he inherited from his father, to arrange, classify and collect statistics about all sorts of things. At the same time he was deeply interested in problems of biology, particularly those having to do with inheritance. His interest in this direction crystallized into definite activity at about the time when his cousin Charles Darwin was elaborating his theory of heredity, which was called pangenesis. Galton instantly realized that this conception of the physiology of the hereditary process was essentially statistical in character, and that statistical methods were demanded to test and broaden it. Upon this work he therefore embarked with the vigor and ardent enthusiasm which characterized all of his scientific work. His results found expression in a series of memoirs and books which have become classics in biological science. Of these the most important is perhaps "Natural Inheritance," since in it are brought to a focus a number of different lines of work which engaged Galton's thought and energy for many years. In this book the attempt is made for the first time to determine, on a statistical basis, the degree of resemblance, in respect of bodily, mental, and temperamental traits which obtains between relatives of different degrees. Previously no attempt had been made to measure precisely these resemblances, which were, of course, a matter of common observation, though not of precise definition, to everyone.

In order to make the desired analysis of this problem it was necessary for Galton to devise new methods of dealing with statistics. The general mathematical foundations of statistical science had, to be sure, been laid by Laplace and Gauss, and some progress in the application of these methods had been made by the man usually regarded as the founder of modern statistics, Quetelet, who was an honorary member of this Association in its early days. But none of these men had dealt specifically with the measurement of what are now known as correlated variations. From Galton's point of viewing the problem of heredity such a measure was an absolute necessity. He, therefore, devised one. It was not altogether a perfect one but was practically usable, and led very shortly to the application, by other hands, of an old theorem of Bravais, which furnished the entirely adequate measure which Galton had sought.

To the end of his life Sir Francis Galton retained his interest in the science of biometry, of which he may truly be said to have been the founder. His keenness of interest served in great part as the primal inspiration and stimulus which led two other distinguished English workers to enter this field and begin to rear the superstructure on the foundation already laid. I refer to Prof. Karl Pearson of University College and the late Prof. W. F. R. Weldon, whose untimely death took from English science one of her most brilliant intellects and charming personalities. To Professor Pearson belongs the very great credit of developing adequate and general mathematical methods for the analysis of biological statistics. Statistical mathematics in the main fall within the realm of the calculus of probability. The foundations of that calculus were laid by Laplace and Gauss, as has already been pointed out. Since their day the most notable fundamental advance in the mathematical theory of probability has, in the writer's judgment, been due to the genius of Karl Pearson. Until he began his work all statisticians, astronomers, and physicists who had anything to do with the theory of probability, either from the standpoint of statistics or that of the theory of errors of observation, had been content to use the so-called ''normal'' curve of errors to describe the distribution of chance determined events. One of the characteristics of this curve is that it is symmetrical. According to it events above the mean are as likely to happen as events below the mean. Observed statistics of natural phenomena were found, as a matter of fact, to give in many cases asymmetrical distributions. Indeed, some of the very examples used in the text-books to illustrate the normal curve do not accord with it when tested by an accurate measure of goodness of fit (for which extremely valuable instrument of statistical research we are again indebted to Pearson). That convenient scape-goat ''chance'' was appealed to, however, to account for the discrepancies. Starting from the sound position that the facts of nature are of more importance than any theory, even though it be one beautiful enough to excite worship,* Pearson in three classic memoirs, in the series of ''Mathematical Contributions to the Theory of Evolution,'' developed a theory of skew frequency curves, and skew correlation, which took account of the asymmetry so frequently seen in chance determined phenomena. This system of skew frequency curves has had the test now of nearly twenty years' usage. Every attempt at destructive criticism which has been aimed against it has failed. None of the substitutes, some of which have been proposed by very eminent mathematicians, have shown any approach to the generality and elegance of these curves.

Few biologists have an adequate conception of the extent to which biometry is indebted to Prof. Karl Pearson. If, as has been maintained, every real advance in science depends upon the discovery and perfection of a new technique, then, for whatever advance in biology may come through biometry, the debt to that distinguished investigator will be large for many years to come.

With so much by way of historical orientation we may now turn to a consideration of the actual concrete results which the biometrical method of attacking biological problems has yielded. Here, of course, only a few of the more important and outstanding points can be touched upon. Anything like a complete review of the pertinent literature is clearly out of the question in view of the fact that the statistical method is literally coming to permeate all fields of biological research. Starting, as we have seen, with the application to the problems of heredity, biometric methods first were taken into other phases of the evolution problem, as, for example, variation simple and correlated, natural selection, homogamy, etc. From these fields to comparative anatomy and taxonomy was an easy step. The whole field of experimental embryology and *Entwicklungsmechanik* clearly offered splendid opportunities for the biological statistician to bring his new technique into operation, which he was not slow to realize. In the field of animal behavior and comparative psychology statistical methods have taken a prominent place in the investigator's armamentarium, doubtless the more readily because of the considerable development of *Psychophysik* in the domain of human psychology. Finally, and most recently, biometric methods are making a place for themselves in cytology.

---

* Galton always insisted that the Greeks would have deified the "law of error" had they known of it.

Out of such a wealth of material it is difficult to single particular researches for special mention. Instead let us try to examine in the brief time at our disposal some of the more noteworthy results of biometry in certain broad general lines. Foremost stands the subject of heredity. Here a curiously complicated condition of affairs has existed. At just the time (1900) when biometric researches in this field were being most actively prosecuted by Pearson and his co-workers came the rediscovery of the work of Gregor Mendel, and the laws of inheritance which that work established. It was at once perceived that these results gave a fundamentally new conception of the physiology of the hereditary process. This was, in certain respects, completely at variance with the concept of heredity developed by Darwin and Galton, which had furnished the leading idea in the investigations of Pearson. In the judgment of the great majority of the workers in this field, the newer Mendelian principles rendered the older methods of research essentially futile. It appeared that they were not adapted to get at the real kernel of the problem. This opinion has gained strength with the passing of time. It led, however, to the premature and unfortunate conclusion that, because biometric methods as applied to the problem of heredity in one particular case and one particular way did not lead to fruitful results, therefore these methods were of no use under any circumstances in dealing with this problem. Happily this view found an immediate corrective in the investigations of the distinguished Danish botanist, Prof. W. Johannsen. His *"Elemente der exakten Erblichkeitslehre,"* of which a new and greatly enlarged edition is just off the press, gave a clear demonstration that biometric methods, when guided by sound biological ideas, were not only a great aid in the elucidation of the problems of heredity, but were in fact essentials. These methods enabled Johannsen to demonstrate with convincing clearness the fundamental distinction between phenotypic and genotypic variations; those, on the one hand, which have to do solely with the soma and not at all with inheritance, and those, on the other hand, which are inherited. This is a most fundamental concept, upon which all modern progress in the study of this great problem is based. Furthermore by the aid of biometric methods Johannsen was able to show the real nature and mode of operation of selection, artificial and natural. These results mark a real debt of biology to statistics.

It may fairly be said that the results which have been obtained by the application of statistical methods to biological problems in the field of *Entwicklungsmechanik* stands second in importance only to those in the fields of evolution and heredity. Biometry has furnished a valuable adjunct to the experimental method in the analysis of many problems of morphogenesis. In particular attention should be called to studies on growth, and also to studies on the general tectonic of the organism. By the application of appropriate biometric methods two fundamental laws of growth of wide generality in both the plant and animal kingdoms have been established. The first of these relates to absolute growth increments, and states that as an organism increases in size the absolute increment per unit of time becomes progressively smaller, in accordance with a logarithmic curve. In other words, the size which an organism has attained, after having grown a given time, is a logarithmic function, simple or complex as the case may be, of the time that it has been growing. The second law of growth, which, like the first, appears to be of wide generality, relates to the variability of the growing organisms, and states that relative variability tends to decrease progressively as growth continues. The problem of growth has been particularly studied from the biometrical standpoint, for some years past, by Dr. H. H. Donaldson and Dr. Shinkishi Hatai, of the Wistar Institute of Anatomy, using chiefly the white rat as material for investigation. The pioneer researches in this field were those of Prof. Charles Sedgwick Minot of Harvard Medical School, on guinea pigs. Other workers have studied with similar results the same problem in a number of plants.

Directly connected with the studies on growth has been the quantitative investigation of the principles of morphological organization. This I have referred to above as the problem of the general tectonic of the living body. This problem is one which has by no means as yet received the attention which is its due. It is one on which little real progress can be made except by the use of biometric methods.

Turning again to the general field of organic evolution a very important result of the statistical method has been the proof which it has given of the existence of assortative mating in the breeding of animals under natural conditions, and the measurement of the degree of this assortative mating or homogamy in particular cases. The significance of homogamy as a *vera causa* of evolution was first adequately emphasized by the post-Darwinians, and particularly by Romanes. He pointed out that: "Isolation takes rank with Heredity and Variability as one of the most fundamental principles of organic evolution. For, if these other two principles be granted, the whole theory of descent resolves itself into an inquiry touching the causes, forms and degrees of Homogamy." When this was written (1897) there was practically no definite and precise knowledge in even a single instance "touching the causes, forms and degrees of Homogamy." Since that time, as a result of the application of biometric methods to this problem, the degree of homogamy has been precisely measured in a number of cases, in some of which even its existence was not before suspected. Pearson and his co-workers have studied assortative mating in man. They have been able to show that for a wide variety of characters there is a definite and statistically significant degree of resemblance between husband and wife. Going to the other end of the evolutionary scale it has been conclusively and independently demonstrated by Prof. H. S. Jennings of Johns Hopkins University, and the present speaker, that in the conjugation of Paramecium there is a high degree of assortative mating; "like pairs with like" to a marked degree. A rich harvest awaits further tilling of this field. Clear as is the significance of homogamy as a factor in evolution when viewed from the older standpoint emphasized by Romanes, the whole matter gains added significance in the light of the newer studies of heredity. The important investigations of Prof. E. M. East of Harvard University on heterozygosis clearly indicate, as a collateral issue, the significance of studying the degree of assortative mating which occurs in the reproduction of wild forms, for homogamy tends automatically to produce homozygosis. It may be said that we are not likely to gain any adequate conception or measure of the significance of heterozygosis as a factor in organic evolution until we have in hand the results of more extended investigations than have yet been made regarding the frequency of occurrence and degree of assortative mating in nature.

We may now turn in conclusion to the more general and philosophical aspects of the relation of statistical science to biology. As I have elsewhere pointed out, "the real purpose of biometry is the general quantification of biology. Its fundamental viewpoint is that without a study of the quantitative relations of biological phenomena in the widest sense it will never be possible to arrive at a full and adequate knowledge of those phenomena. This viewpoint insists that a description which says nothing about the magnitude of the thing described is not complete, but, on the contrary, lacks an element of primary importance. It insists, also, that an experiment which takes no account of the probable error of the results reached is inadequate and as likely as not to lead to incorrect conclusions."

Statistical science has brought to biology three fundamentally important things which it had previously lacked. These are: first, a method of describing a *group* of individuals in terms, not of its component individuals, but in terms of its (the group's) own attributes and qualities; second, the concept of ''probable error,'' which makes possible an estimate of the probable accuracy of a series of observations; and third, a method of measuring the degree of association or correlation between the variations in a series of characters or events. Time is lacking to develop each of these points, nor is it necessary, for they have been discussed in detail elsewhere.

In what has preceded I have tried to show that by turning to statistical science for aid the biologist has greatly augmented his powers of analysis in the domain of his own particular problems. While the branch of science, which has been called into being by this coalition, is yet too young to have shown its full capabilities, yet I think its achievements have been sufficient in quality and amount to justify the belief that its position is secure and its promise bright. Biometry seems destined to become a permanent and important branch, at once of biological investigation and of statistical inquiry.